

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221315092>

Recognizing Human Actions Using Silhouette-based HMM

Conference Paper · September 2009

DOI: 10.1109/AVSS.2009.46 · Source: DBLP

CITATIONS

53

READS

292

5 authors, including:



Carlos Orrite

University of Zaragoza

89 PUBLICATIONS 1,296 CITATIONS

[SEE PROFILE](#)



Elías Herrero Jaraba

University of Zaragoza

39 PUBLICATIONS 384 CITATIONS

[SEE PROFILE](#)



Hossein Ragheb

Renishaw plc

70 PUBLICATIONS 842 CITATIONS

[SEE PROFILE](#)



Sergio Velastin

University Carlos III de Madrid

300 PUBLICATIONS 8,730 CITATIONS

[SEE PROFILE](#)

Recognizing Human Actions using Silhouette-based HMM

Francisco Martínez-Contreras, Carlos Orrite-Uruñuela, Elías Herrero-Jaraba
CVLab, Aragon Institute for Engineering Research
University of Zaragoza, SPAIN
franjar@unizar.es, corrite@unizar.es, jeliass@unizar.es

Hossein Ragheb, Sergio A. Velastin
Digital Imaging Research Centre
Kingston University, UK
sergio.velastin@kingston.ac.uk

Abstract—This paper addresses the problem of silhouette-based human action modelling and recognition, specially when the number of action samples is scarce. The first step of the proposed system is the 2D modelling of human actions based on motion templates, by means of Motion History Images (MHI). These templates are projected into a new subspace using the Kohonen Self Organizing feature Map (SOM), which groups viewpoint (spatial) and movement (temporal) in a principal manifold, and models the high dimensional space of static templates. The next step is based on the Hidden Markov Models (HMM) in order to track the map behavior on the temporal sequences of MHI. Every new MHI pattern is compared with the features map obtained during the training. The index of the winner neuron is considered as discrete observation for the HMM. If the number of samples is not enough, a sampling technique, the Sampling Importance Resampling (SIR) algorithm, is applied in order to increase the number of observations for the HMM. Finally, temporal pattern recognition is accomplished by a Maximum Likelihood (ML) classifier. We demonstrate this approach on two publicly-available dataset: one based on real actors and another one based on virtual actors.

Keywords—Activity Recognition, MHI, SOM, HMM, SIR, ViHaSi, MuHaVi

I. INTRODUCTION

In recent years, human activity recognition has experienced an increased relevancy due to the interest in the development of new autonomous video surveillance systems and the importance of ambient intelligent applications in modern life, such as supportive home environment for elderly and disabled people. Either because of the large number of cameras deployed in public spaces or the privacy aspects of their use in private spaces (homes), it is not possible to manually monitor the video stream coming from these cameras. For this reason, it is important to develop computer vision systems able to process video data and classify and recognize human behaviour.

This paper describes a new method to deal with the temporal features needed for the detection of human actions using a Kohonen's Self-Organized Map (SOM) to model temporal templates in the lower dimensional space formed by the M neurons whose characteristics are tracked in time by means of a HMM to carry out the action recognition process. In order to train an action-specific SOM many samples are needed so as to achieve stable and accurate

performance [1]. When this is not possible, a possible solution is to train a single SOM with all the actions. This implies an additional procedure to track different neuronal activation on a temporal basis so as to classify different actions using, for example, action-specific HMMs. Again, when the number of action samples is not enough, the same problem arises. To manage it we propose a SOM sampling technique to increase the number of examples.

II. PREVIOUS WORK

Human action recognition and pose recovery have been studied extensively in recent years, see [2] and [3] for a survey. In this paper, a brief overview of those methods which are more related to our proposal is given.

Silhouette based methods have received significant attention lately. Motion Energy Images (MEI) and Motion History Images (MHI) were introduced by Davis and Bobick [4], to capture, on the one hand, where motion occurred and, on the other, the history of motion occurrences in the image. This approach works effectively under the assumption that the viewpoint is relatively fixed. To overcome this limitation, the authors propose to use multiple cameras. Weinland et al. [5], extend the MHI concept introducing motion history volumes to achieve a free-viewpoint representation for human actions in the case of multiple calibrated video cameras. This method is limited since it requires training and testing on multiple cameras. More recently [6], the same authors have proposed a framework to model actions using three dimensional occupancy grids, built from multiple viewpoints, in an exemplar-based HMM. Lv and Nevatia [7] exploit a similar idea, where each action is modelled as a series of synthetic 2D human poses rendered from a wide range of viewpoints, instead of using 3D explicitly. The constraints on transition of the synthetic poses are represented by a graph model called Action Net. Given the input, silhouette matching between the input frames and the key poses is performed first using an enhanced Pyramid Match Kernel algorithm. The best matched sequence of actions is then tracked using the Viterbi algorithm.

The use of SOM networks for human activity recognition has not received much attention in the past. Johnson and Hogg [8], describe object movement as positions and velocities in the image plane. A statistical model of object trajec-

tories is built with two competitive learning networks. Hu et al. [9], introduce a hierarchical self-organizing approach for learning the patterns of motion trajectories that has smaller scale and faster learning speed. Owens and Hunter [10], apply a SOM feature map to find flow vector distribution patterns. These patterns are used to determine whether a point on a trajectory is normal or abnormal. The use of the SOM in these works is restricted to learning trajectories, rather than modelling the human motion from different point of views.

In previous work [11] we use the same structure of temporal templates, MHI, training a SOM network for each specific action. Action recognition is accomplished by a Maximum Likelihood (ML) classifier over all specific-action SOMs.

Recently, HMM has been successfully applied to the task of human action recognition. Related our work include that of Kellokumpu et al. [12] which describes actions as a continuous sequence of discrete postures derived from an invariant descriptor. They extract the contour from human silhouettes and calculate affine invariant Fourier descriptors from this contour. In order to classify the posture, these descriptors are used as the inputs of a radial basis SVM. They use HMM to model different activities and calculate the probabilities of the activities based on the posture sequence from SVM. Wang and Suter, in [13], propose the following method: First, features are extracted from image sequences, with two alternatives: silhouette and a distance transformed silhouette; the second step is to find a low-dimensional feature representation embedded in high-dimensional image data, and thirdly, analyze and recognize the motions using a continuous HMM. Kale et al., in [14], propose an continuous HMM-based approach to represent and recognize gait. Their approach involves deriving a low dimensional observation sequence (the width of the silhouette of the body) during a gait cycle. Learning is achieved by training an HMM for each person over several gait cycles.

A fast and effective method for classifying actions based on Fourier descriptors of silhouettes has been described by Rageb et al in [15].

III. PROPOSAL OVERVIEW

The aim of this paper is to present a new approach to the recognition of human action when the number of action samples is small. For this purpose, the authors use MHIs, as introduced in [11], as temporal features. MHIs capture spacial and temporal information of the motion in images, encoding how recently this motion occurred. A real time system cannot work with the high dimensionality of the information stored in the MHIs needed for the recognition of the actions set. For this reason, all these 2D motion templates are projected into a new subspace by means of a single SOM [16] with all the actions.

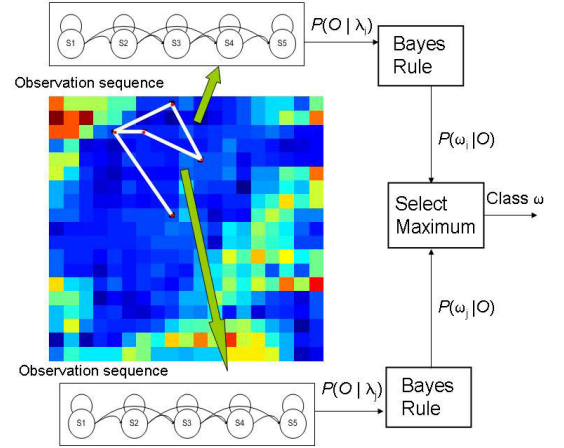


Figure 1. Maximum likelihood classifier for temporal templates based on SOM and HMMs, where the map of 15x15 neurons is depicted by its U-matrix, and the lines are activation trajectories with the temporal sequence of patterns

Once the temporal templates have been mapped into the SOM, an action specific HMM is trained in order to track the map behavior on the temporal sequences of static templates. Again, when the number of action samples is not large enough, it is necessary to use some sampling technique over the SOM outputs to increase the number of samples. Two different kinds of observations can be obtained from the SOM: the index of winner neurons, as discrete observation, and the quantization error distribution, as continuous observation.

Finally, temporal pattern recognition is accomplished by a Maximum Likelihood (ML) classifier. Fig. 1 shows the structure of the ML classifier, where a map of 15x15 neurons is depicted by its U-matrix, and the lines represent the activation trajectory for a particular temporal sequence of patterns. The U-matrix visualizes distances between neighbouring map units, and helps to see the cluster structure of the map. Different HMMs receive this observation sequence as input for estimating the probability of class membership.

Following Bayes rule, the posterior is given by:

$$p(\omega_i | O) = \frac{p(\omega_i) \cdot p(O | \omega_i)}{p(O)} \quad (1)$$

In the equation 1, $p(\omega_i)$ is the priori likelihood, where ω_i is the class i . The conditional term, $p(O | \omega_i) \approx p(O | \lambda_i)$, with λ_i , the parameters of the HMM_i , can be calculated by the forward-backward algorithm and the probability of the observation $p(O)$ is the same for all classes so, it is not taken into account to calculate the maximum a posteriori.

IV. HUMAN ACTION MODELING BASED ON MOTION TEMPLATES

Motion templates are based on the observation that in video sequences a human action generates a space-time

shape in the space-time volume. MHIs were introduced by Bobick and Davis [4], to capture motion information in images, encoding how recently motion occurred at a pixel. Consider $D(x,y,t)$ a binary image sequence of motion regions, where $D(x,y,t) = 1$ indicates that some motion was present at time t and at location (x, y) . Then, each pixel intensity of the MHI is a function of temporal history of motion at that point, occurring during a fixed duration τ (where $1 \leq \tau \leq N$, for a sequence of length N frames). The result is a scalar-valued image where more recently moving pixels are brighter.

$$MHI_t(x, y) = \begin{cases} \tau & \text{if } D(x,y,t) = 1 \\ \max(0, MHI_{t-1}(x, y) - 1) & \text{otherwise} \end{cases} \quad (2)$$

Every MHI implicitly captures dynamic information of the human figure at any time such as the global body motion or the relative motion of the limbs to the body, as well as spatial information, such as the location and orientation of the torso and limbs or the aspect ratio of different body. Each MHI can be seen as a "piece" of an action and the link of successive pieces will constitute the proper action.

These motion templates strongly depend on image location, scale or view-point. As the goal is to recognize human body activities from different persons that are free to move in space, with different orientations and sizes, some form of normalization step must be carried out. Location and scale dependencies are removed by centering the MHI with respect to the center of mass of the detected object and scaling with respect to the minimum enclosing rectangle [11].

To address dependence on the camera point of view, one possibility would be to generate a 3D model of the body as [5]. However, this approach, using many cameras, each fully calibrated, is not easy to obtain for real sequences. The authors [11] have proposed a novel alternative approach in which 2D motion templates, viewpoint-independent, are obtained projecting the MHI obtained before, for all available viewpoints, into a new subspace by means of the SOM [16].

The main problem with the previous temporal representation is that it involves dealing with a very high dimensional space. An alternative to alleviate this problem consists in mapping temporal features into a lower dimensional space. The simplest way to reduce dimensionality is via Principal Component Analysis (PCA) which assumes that the data lies on a linear subspace. Except in very special cases, data does not lie on a linear subspace, thus requiring methods that can learn the intrinsic geometry of the manifold from a large number of samples [3]. Nonlinear dimensionality reduction techniques allow for representation of data points based on their proximity to each other on nonlinear manifolds.

The SOM is an unsupervised neural network mapping a set of n -dimensional vectors to a two-dimensional topographic map displaying in such a way that similar data items

are located close to each other on the map. The SOM is similar to a k-means clustering algorithm, extending it by providing a topology preserving mapping and placing similar objects in neighboring clusters.

During the training process a map is built and the neural network organizes itself, using a competitive process. The network must be given a large number of input vectors, as much as possible representing the kind of vectors that are expected. The basic idea of a SOM is simple: every neuron i of the map is associated with an n dimensional codebook vector $m_i = (m_{i_1}, \dots, m_{i_n})$, where n is the dimension of the input patterns. The neurons of the map are connected to adjacent neurons by a neighbourhood relation, which defines the distribution of the map over the data space. The network is trained by finding the codebook vector that is most similar to an input training vector. This codebook vector and its neighbours are then updated so as to render them more similar to the input vector.

Fig. 2 shows a representation of the codebook of the trained SOM of size 15×15 for a set of actions. Each neuron is represented in the map by the pattern learned during the training.

During the test stage, for each new MHI input pattern, a Quantization Error Distribution (Q-error) is obtained, calculating the Euclidean distance between this input vector and weight vectors of the neurons of the map. The winner neuron, is which whose weight vector lies closest to the input vector.

Since our purpose is to use the SOM behavior along time to model the temporal features of an action, we have to measure dissimilarity between map behaviors in response to two different patterns. Two kind of observations provided by the SOM are observed: the index of the winner neurons and the Q-error distribution.

V. RECOGNITION USING HMM

Formally, the parameters of the Markov model, notated as $\lambda = \{A, B, \pi\}$, are specified in our particular problem as follows [17]:

- 1) N states, i.e. $S = S_1, \dots, S_N$
- 2) The initial state probability distribution, $\pi = \{\pi_i\}$ where $\pi_i = Pr(q_1 = S_i)$, $1 \leq i \leq N$
- 3) State transition matrix, $A = a_{ij}$, where the transition probability a_{ij} represents the frequency of transiting from state i to state j calculated as: $a_{ij} = Pr(q_{t+1} = S_j \mid q_t = S_i)$
- 4) Observation probability distribution, $B = b_j(o_t)$, where $b_j(o_t) = Pr(o_t \mid q_t = S_j)$ and o_t the observation vector

The problem of estimating the parameters λ of a HMM, given a sequence of observations $O = \{o_1, \dots, o_T\}$, can be approached as an ML problem. To solve this we have the Baum-Welch algorithm, which is an EM procedure, estimating the parameters of a HMM in an iterative procedure.

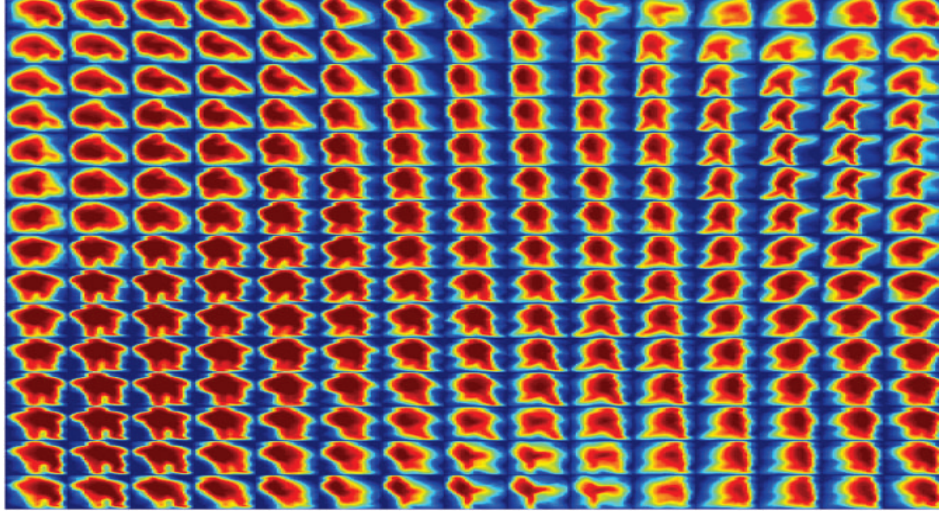


Figure 2. Trained SOM for the 8 actions of the MuHaVi-MAS database

A. Discrete Inputs

Given a set of Q-error maps, Qe_t with $t = \{1, \dots, T\}$, a sequence of observations is given by the indexes of the winning neurons, known as the Best Matching Units (BMU), i.e., those neurons with the lowest distance error, $O = \{o_1 = BMU_1, \dots, o_T = BMU_T\}$. If the number of temporal sequences to train the HMM is low, then most of these values will never be reached. This constitutes a serious problem when training the HMM, mainly in relation to the observation matrix.

Taking into account that there are neighbor neurons around the winner one with low distance as well, we propose to get more discrete observations, i.e., more indexes, from every Q-error map by using a sampler method.

B. Sampling Importance Resampling (SIR)

Let us consider that every observation is independent to the previous one, which is not true, but it constitutes a naive approach. In this way, we have:

$$p(O) = p(o_1, \dots, o_T) = \prod_{t=1}^T p(o_t) \quad (3)$$

Given a sequence of Q-error maps, the goal is to obtain several samples for every Q-error map, i.e. different indexes $o_t = i$, taking into account the probability of every neuron. This probability can be given as:

$$p(o_t = i) \propto e^{-\lambda \cdot Qe_t(i)^2} \quad (4)$$

where $Qe_t(i)$ is the Q-error of the neuron i at time t and λ is a smooth parameter chosen experimentally. In this way, a distance is transformed into a probability. If the quantization error is null then, the probability is 1. On the other hand, if the distance tents to infinity, the probability is null.

Then, every Q-error map is sampled independently, by for instance a Sampling Importance Resampling (SIR) method [18]. This naive approach has given good results as shown in the next experiments.

VI. EXPERIMENTS

The present approach has been tested using two different datasets, one consisting on virtual actors (ViHaSi) [15], and other one, MuHaVi-MAS with real actors.

A. MuHaVi-MAS Database

The Multicamera Human Action Video (MuHaVi) Manually Annotated Silhouette data (MuHaVi-MAS) [19] is composed of 5 action classes (WalkTurnBack, RunStop, Punch, Kick, ShotgunCollapse) and the authors manually annotated the image frames to generate the corresponding silhouettes of the actors. There are only 2 actors and 2 camera views for these 5 actions. From these action combinations can be obtained 14 simpler actions (collapse right, walk right to left, walk left to right, turn back right...), which may also reorganized in 8 classes (Collapse, Stand up, Kick Right, Punch Right, Guard, Run, Walk, Turn back) where similar actions make a single class. This database, which only has 136 sequences.

The parameters used in order to obtain the motion templates of the MHIs, are $\tau = 10$ frames, and the displacement of the overlapping windows is 1 frame. Due to the scant number of sequences per action, which makes impossible the good training of separate SOMs for each action, we have decided to train only a single SOM, of size 15×15 , for the whole set of actions, and apply the "leave one-out" technique over this training. That is, we train with 135 sequences, and test with the remaining one.

In order to evaluate the performance of the HMM as neuron tracker, several HMMs, with different numbers of

1	2	3	4	5	6	7	8
100	100	100	100	93.75	93.75	100	100

Table I
SUCCESSFUL RATE, IN PERCENT, AGAINST ACTION

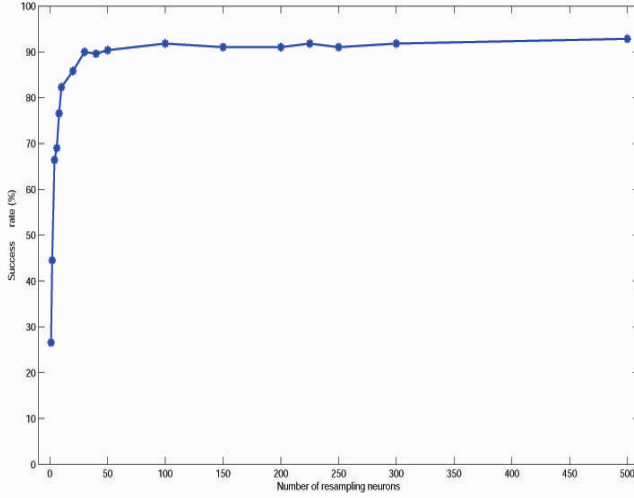


Figure 3. Successful rate as a function of the number of resampling neurons

hidden states (from 2 to 11) have been tested, using different number of neurons as discrete input. The best results, for diverse values of HMM states and number of resampling neurons, are shown in the Table I. The overall recognition rate is 98.4375%.

The confusion matrix using the "leave one-out" cross validation approach showing the number of test sequences classified in each class is shown in Table II. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The number of sequences of each action is shown in the Table III.

In the Figure 3 can be seen the evolution of the overall rate from a 2-states HMM depending on the number of resampling neurons. It is clear to observe that the resampling technique improve the results for a given HMM.

B. ViHaSi Database

Virtual Human Action Silhouette (ViHaSi) data [20], provides a large body of synthetic video data generated for the purpose of evaluating different algorithms on human action recognition which are based on silhouettes. This dataset exhibits as an interesting property that different fitting clothes are worn by the actors. The data consists of 20 action classes, 9 actors and up to 40 synchronized perspective camera views split into two sets of 20 cameras. The test was carried out by the "leave-one-out" cross validation over all 20 actions, all actors, and 12 camera views. Each action

action	1	2	3	4	5	6	7	8
1	16	0	0	0	0	0	0	0
2	0	12	0	0	0	0	0	0
3	0	0	16	0	0	0	0	0
4	0	0	0	16	0	0	0	0
5	0	0	0	0	30	0	2	0
6	0	0	1	0	0	15	0	0
7	0	0	0	0	0	0	16	0
8	0	0	0	0	0	0	0	12

Table II
CONFUSION MATRIX

action	1	2	3	4	5	6	7	8
number of sequences	16	12	16	16	32	16	16	12

Table III
NUMBER OF SEQUENCES FOR EACH ACTION

sequence, over 2640, is formed by 5 MHIs, and only one SOM, of size 29 x 20 is trained for the whole set of actions. Different number of states of HMM and different number of resampling neurons are tested. The average recognition rates per action are shown in the Table IV. The overall recognition rate is 99.92%.

VII. CONCLUSION

In this paper a new method for resolving the problem of action recognition from temporal series is presented. The proposal is based on the use of temporal templates, MHI, to capture both spatial and dynamic information about the human silhouette, mapping these features into a lower dimensional space by means of a SOM. The trajectory of the temporal sequence of an action over the map is tracked, obtaining the winner neurons and their quantization error distribution, which are used as input of the several discrete HMM in order to recognize the action. Due to the scarce number of data, it is not sufficient with the SOM observations to train the HMM.

One of the proposal of this approach is the use of the resampling techniques as SIR algorithm, widely used in the particle filter, to increase the number of training sequences, useful when the number of action samples is not enough. The results shown in the Figure 3 prove that the use of this

1	2	3	4	5	6	7	8	9	10
100	100	100	100	100	100	99.24	100	100	100
11	12	13	14	15	16	17	18	19	20
100	100	99.24	100	100	100	100	100	100	100

Table IV
SUCCESSFUL RATE, IN PERCENT, PER ACTION

algorithm significantly improves recognition rates through the use of resampling. Test average recognition rates given in a real dataset are very promising considering the complexity of the test and the small amount of data used for training. Comparing with other authors with other databases, as Irani et al. in [21], the results obtained in MuHaVi-MAS are at least comparable with the obtained by them. The results in the ViHaSi dataset improve the obtained ones by the authors in [11], from 98.48% to 99.92%. It is important that the action recognition community increases the use of a reference dataset such as MuHaVi so that it becomes easier to compare results. The overall processing time in Matlab takes less than 12 seconds on a Pentium 4, 3.4 GHz.

As future work we want to evaluate the performance of HMM using as continuous input Q-error, for which it is essential to obtain a good initialization of the parameters of the Markov models.

ACKNOWLEDGEMENTS

This work is partially supported by grant TIN-2006-11044 (MEyC) and FEDER. The Kingston team was supported by the EPSRC's REASON research grant. Dr. Orrite received a Grant by DGA(CONAID) and CAI.

REFERENCES

- [1] P. Zegers and M. Sundareshan, "Systematic testing of generalization level during training in regression-type learning scenarios," *IEEE Int Joint Conf on Neural Networks*, vol. 4, pp. 2807–2812, 2004.
- [2] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 34, pp. 334–352, 2004.
- [3] P. K. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 18, no. 11, pp. 1473–1488, 2008.
- [4] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001.
- [5] D. Weinland, R. Ronfar, and E. Boyer, "Free viewpoint action recognition using motion history volumes," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 249–257, 2006.
- [6] D. Weinland, E. Boyer, and R. Ronfard, "Action recognition from arbitrary views using 3d exemplars," in *Proceedings of the International Conference on Computer Vision*, 2007, pp. 1–7.
- [7] F. Lv and R. Nevatia, "Single view human action recognition using key pose matching and viterbi path searching," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, pp. 1–8, 2007.
- [8] N. Johnson and D. Hogg, "Learning the distribution of object trajectories for event recognition," *Image and Vision Computing*, vol. 14, pp. 583–592, 1996.
- [9] W. Hu, D. Xie, and T. Tan, "A hierarchical self-organizing approach for learning the patterns of motion trajectories," *IEEE Transactions on Neural Networks*, vol. 15, pp. 135–144, 2004.
- [10] J. Owens and A. Hunter, "Application of the self-organizing map to trajectory classification," in *VS '00: Proceedings of the Third IEEE International Workshop on Visual Surveillance*, 2000, pp. 77–83.
- [11] C. Orrite, F. Martinez-Contreras, E. Herrero, H. Ragheb, and S. A. Velastin, "Independent viewpoint silhouette-based human action modelling and recognition," in *Proceedings of the International Workshop on Machine Learning for Vision-based Motion Analysis (MLVMA'08)*, 2008.
- [12] V. Kellokumpu, M. Pietikäinen, and J. Heikkilä, "Human activity recognition using sequences of postures," in *Proceedings of the IAPR Conference on Machine Vision Applications (IAPR MVA 2005)*, 2005, pp. 570–573.
- [13] L. Wang and D. Suter, "Visual learning and recognition of sequential data manifolds with applications to human movement analysis," *Comput. Vis. Image Underst.*, vol. 110, no. 2, pp. 153–172, 2008.
- [14] A. A. Kale, N. Cuntoor, and V. Krüger, "Gait-based recognition of humans using continuous hmms," in *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, 2002, p. 336.
- [15] H. Ragheb, S. A. Velastin, P. Remagnino, and T. Ellis, "Human action recognition using robust power spectrum features," in *ICIP*. IEEE, 2008, pp. 753–756.
- [16] T. Kohonen, *Self-organization and associative memory: 3rd edition*. Springer-Verlag New York, 1989.
- [17] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, 1989, pp. 257–286.
- [18] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.
- [19] D. I. R. C. Kingston-University, "The muhavi-mas database," <http://dipersec.king.ac.uk/MuHAVi-MAS/>, 2008.
- [20] D. I. R. C. Kingston University, "The vihasi database," <http://dipersec.king.ac.uk/VIHASI/>, 2008.
- [21] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2247–2253, 2007.